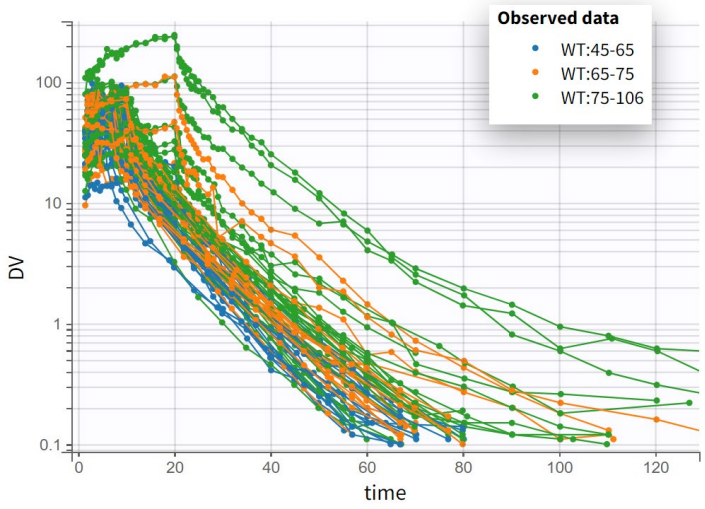
SimulationsPlus

Automated covariate selection: SAMBA and COSSAC algorithms

Claude Magnard

ACoP14 – November 8th 2023

Covariate selection



PARAMETERS	DISTRIBUTIONS	RANDOM EFFECTS	CORRELATION	AGE	BSA	HT	LBM	SEX	WT
		Select: All | None	#1						
Cl	LOGNORMAL	☑	☐	☐	☐	☐	☐	☐	☐
V1	LOGNORMAL	☑	☐	☐	☐	☐	☐	☐	☐
Q2	LOGNORMAL	☑	☐	☐	☐	☐	☐	☐	☐
V2	LOGNORMAL	☑	☐	☐	☐	☐	☐	☐	☐
Q3	LOGNORMAL	☑	☐	☐	☐	☐	☐	☐	☐
V3	LOGNORMAL	☑	☐	☐	☐	☐	☐	☐	☐

?

$(6 \text{ parameters} \times 6 \text{ covariates})^2 > 1000$ possible covariate models

Covariate model building strategies

SCM
(stepwise covariate modeling)
Available in PsN and Monolix

FREM
(full random effects model)
Available in PsN

SCM+
Available in PsN

SAMBA
Available in Monolix

COSSAC
Available in Monolix

LASSO
Available in PsN

Machine Learning
(e.g random forest, neural networks)

WAM
(Wald Approximation Method)

Covariate model building strategies

Most commonly used but requires many runs…

SCM
(stepwise covariate modeling)
Available in PsN and Monolix

FREM
(full random effects model)
Available in PsN

SCM+
Available in PsN

SAMBA
Available in Monolix

COSSAC
Available in Monolix

LASSO
Available in PsN

Machine Learning
(e.g random forest, neural networks)

WAM
(Wald Approximation Method)

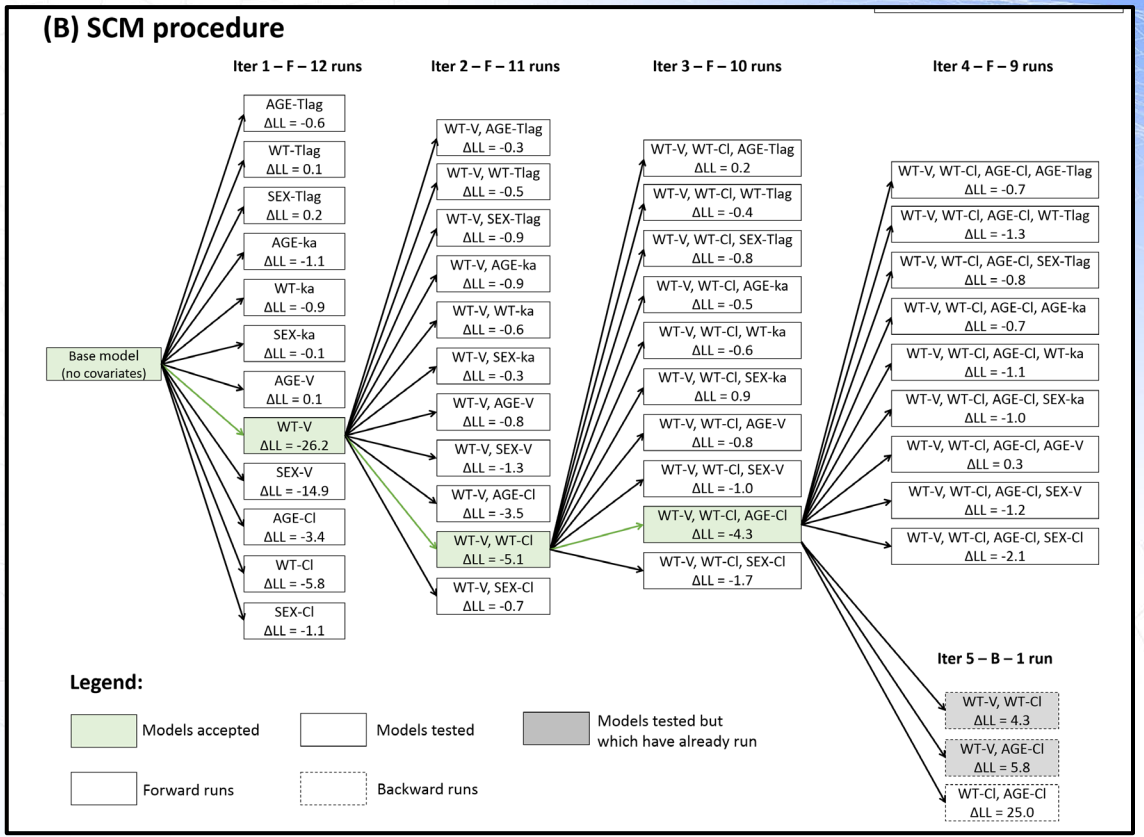
SCM procedure

4 parameters: Tlag, ka, V, Cl and 3 covariates: Age, Weight, Sex

- Test all possible covariate-parameter relationships at each step
- Keep the one that improves LL the most
- Forward until no further addition, then backward

runs ≈ #par x #cov x #relations

takes long to run…



Covariate model building strategies

Most commonly used but requires many runs…

SCM
(stepwise covariate modeling)
Available in PsN and Monolix

FREM
(full random effects model)
Available in PsN

SCM+
Available in PsN

SAMBA
Prague & Lavielle, 2022
Available in Monolix

COSSAC
Ayral *et al*, 2021
Available in Monolix

LASSO
Available in PsN

Machine Learning
(e.g random forest, neural networks)

WAM
(Wald Approximation Method)

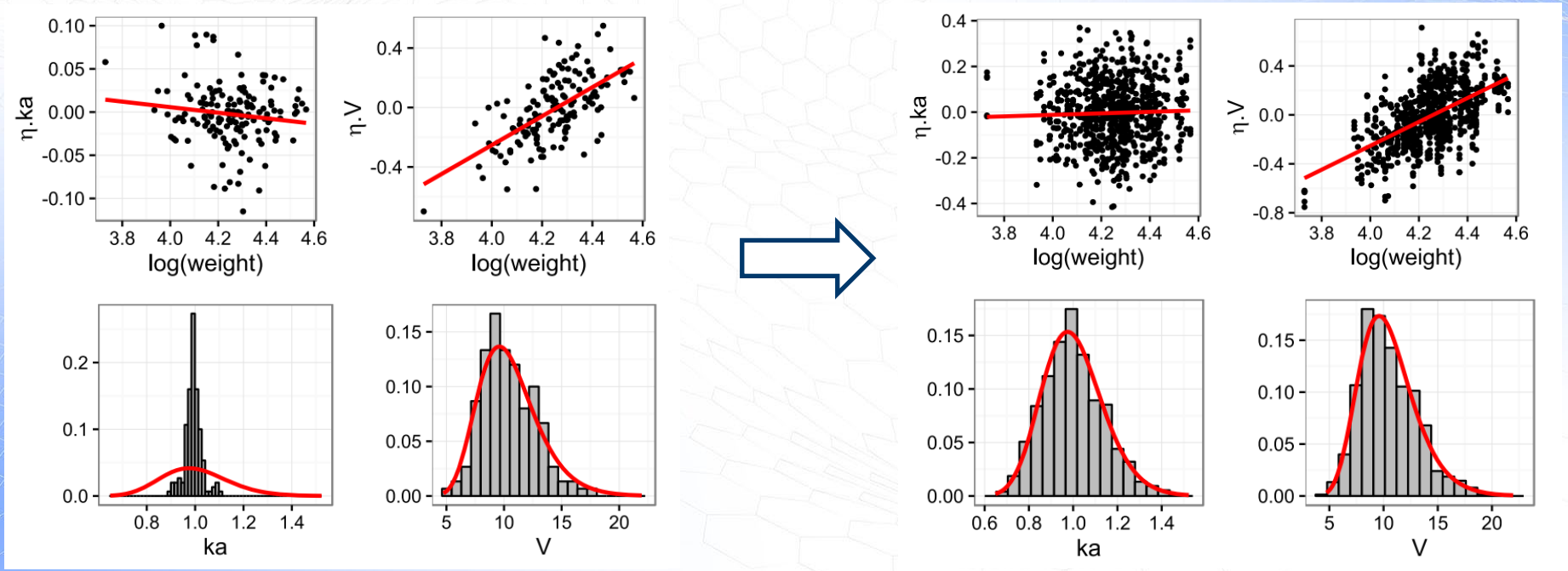
COSSAC

COnditional **S**ampling use for **S**tepwise **A**pproach based on **C**orrelation tests

Ayral (Cellière), G., Si Abdallah, J. F., Magnard, C. & Chauvin, J. A novel method based on unbiased correlations tests for covariate selection in nonlinear mixed effects models: The COSSAC approach. *CPT Pharmacometrics Syst. Pharmacol.* **10**, 318–329 (2021).

COSSAC Key idea

Using samples from the conditional distributions allows to reliably detect correlations between random effects and covariates



COSSAC Key idea

The p-values of the correlation tests can be used to select which covariates to try instead of trying all

Pearson's correlation test and/or ANOVA

eta_Tlag

	COEFF	STATISTICS	P-VALUE
sex		1.14	2.95e-1
age	0.075	0.41	6.84e-1
wt	0.0013	0.0074	9.94e-1

eta_ka

	COEFF	STATISTICS	P-VALUE
sex		0.036	8.5e-1
age	0.0043	0.024	9.81e-1
wt	0.24	1.33	1.94e-1

eta_V

	COEFF	STATISTICS	P-VALUE
sex		18.59	1.61e-4
age	-0.039	-0.21	8.32e-1
wt	0.75	6.24	7.22e-7

eta_Cl

	COEFF	STATISTICS	P-VALUE
sex		0.38	5.43e-1
age	0.32	1.83	7.78e-2
wt	0.35	2.08	4.65e-2

These p-values are unbiased when using samples from the conditional distribution.

Low p-value = significant correlation between samples from conditional distribution and covariates

COSSAC procedure: overview

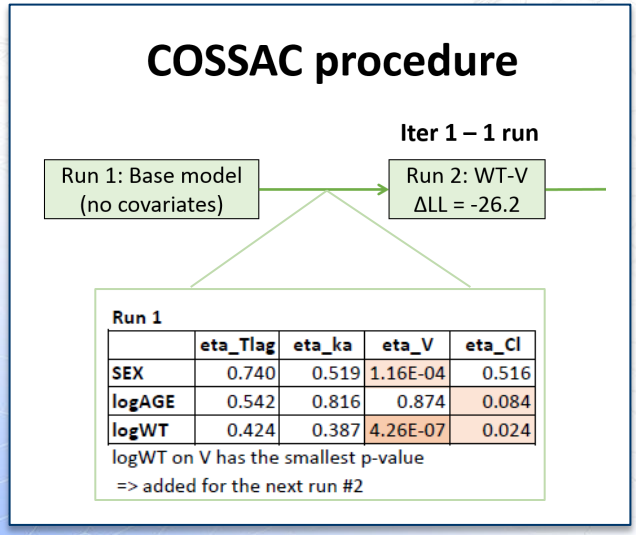
1. Calculate the p-value of the correlation test for each parameter-covariate pair in the current model
2. **Forward step:**
   - Add the most promising (lowest p-value < 0.3) covariate-parameter relationship.
   - If the likelihood improves enough, keep the covariate and add another relationship on top
3. **Backward step:**
   - Remove the least significant (highest p-value > 0.01) already included relationship.
   - If the likelihood worsens too much, put covariate back in the model
4. **Alternate** forward and backward steps

→ Covariates are added one by one (as in SCM, different from SAMBA)

Ayral (Cellière), G., Si Abdallah, J. F., Magnard, C. & Chauvin, J. A novel method based on unbiased correlations tests for covariate selection in nonlinear mixed effects models: The COSSAC approach. *CPT Pharmacometrics Syst. Pharmacol.* **10**, 318–329 (2021).

COSSAC procedure example

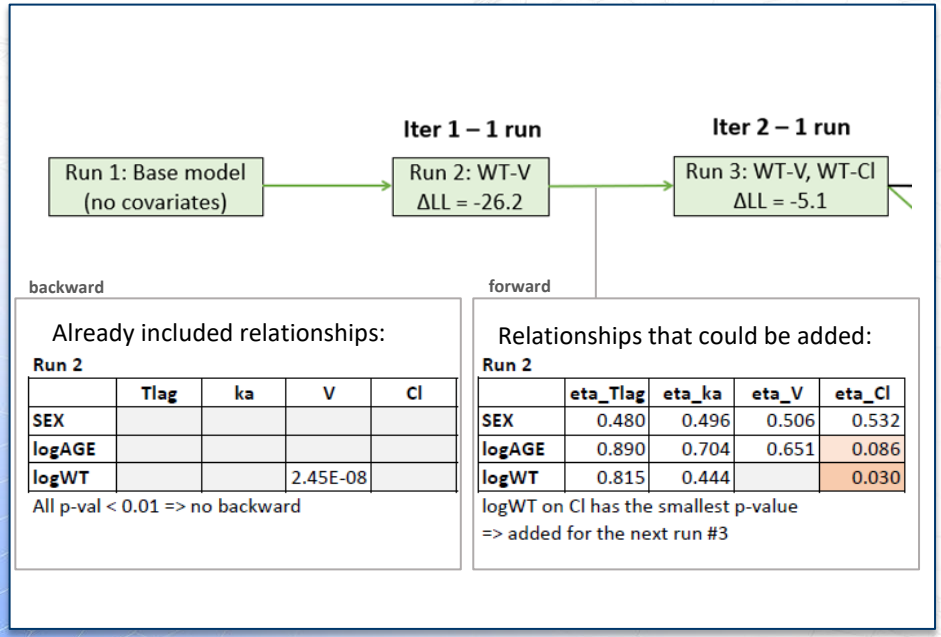
4 parameters: Tlag, ka, V, Cl and 3 covariates: Age, Weight, Sex



Ayral (Cellière), G. et al, *CPT Pharmacometrics Syst. Pharmacol.* (2021)

COSSAC procedure example

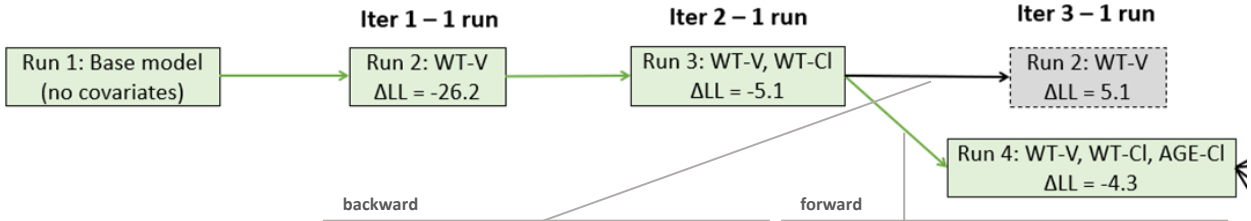
4 parameters: Tlag, ka, V, Cl and 3 covariates: Age, Weight, Sex



Ayral (Cellière), G. et al, *CPT Pharmacometrics Syst. Pharmacol.* (2021)

COSSAC procedure example

4 parameters: Tlag, ka, V, Cl and 3 covariates: Age, Weight, Sex



backward

Already included relationships:

Run 3

	Tlag	ka	V	Cl
SEX				
logAGE				
logWT			4.37E-08	0.025

pval logWT on Cl>0.01 => already tested (run 2)

forward

Relationships that could be added:

Run 3

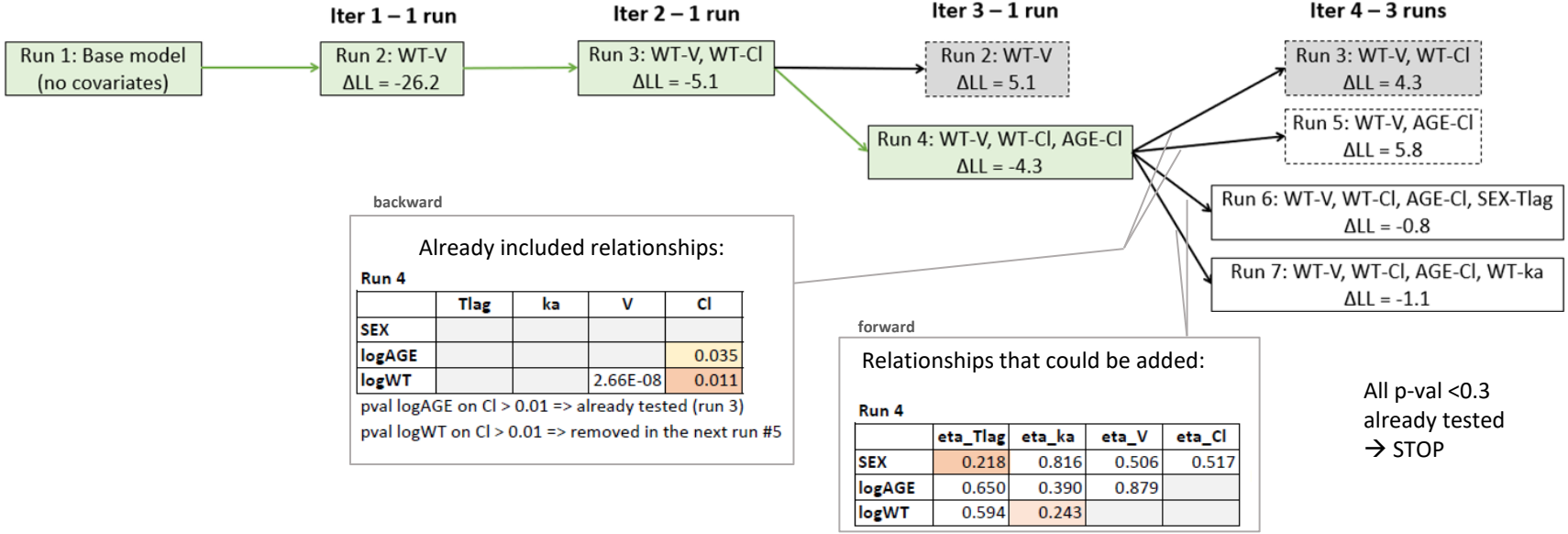
	eta_Tlag	eta_ka	eta_V	eta_Cl
SEX	0.312	0.414	0.511	0.358
logAGE	0.819	0.602	0.748	0.037
logWT	0.742	0.826		

logAGE on Cl has the smallest p-value
=> added for the next run #4

Ayral (Cellière), G. et al, *CPT Pharmacometrics Syst. Pharmacol.* (2021)

COSSAC procedure example

4 parameters: Tlag, ka, V, Cl and 3 covariates: Age, Weight, Sex



	Tlag	ka	V	Cl
SEX				
logAGE				0.035
logWT			2.66E-08	0.011

pval logAGE on Cl > 0.01 => already tested (run 3)

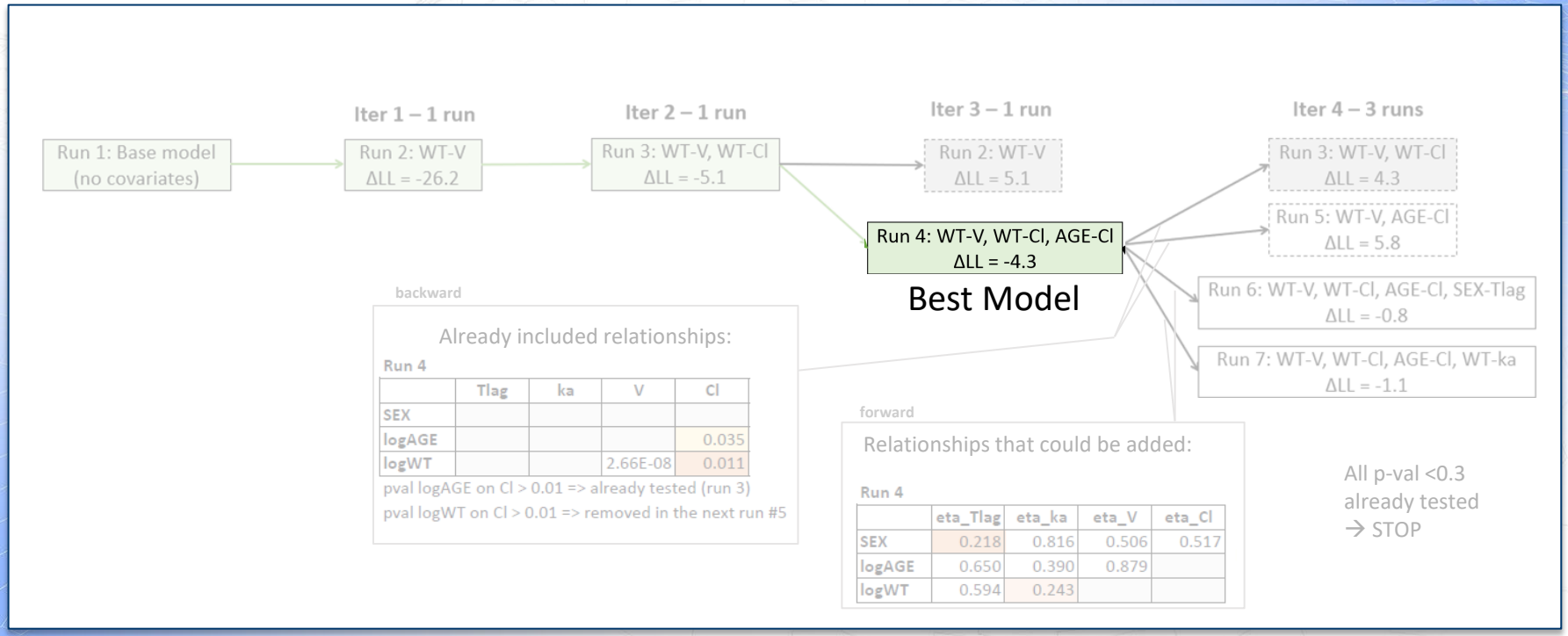
pval logWT on Cl > 0.01 => removed in the next run #5

	eta_Tlag	eta_ka	eta_V	eta_Cl
SEX	0.218	0.816	0.506	0.517
logAGE	0.650	0.390	0.879	
logWT	0.594	0.243		

Ayral (Cellière), G. et al, *CPT Pharmacometrics Syst. Pharmacol.* (2021)

COSSAC procedure example

4 parameters: Tlag, ka, V, Cl and 3 covariates: Age, Weight, Sex

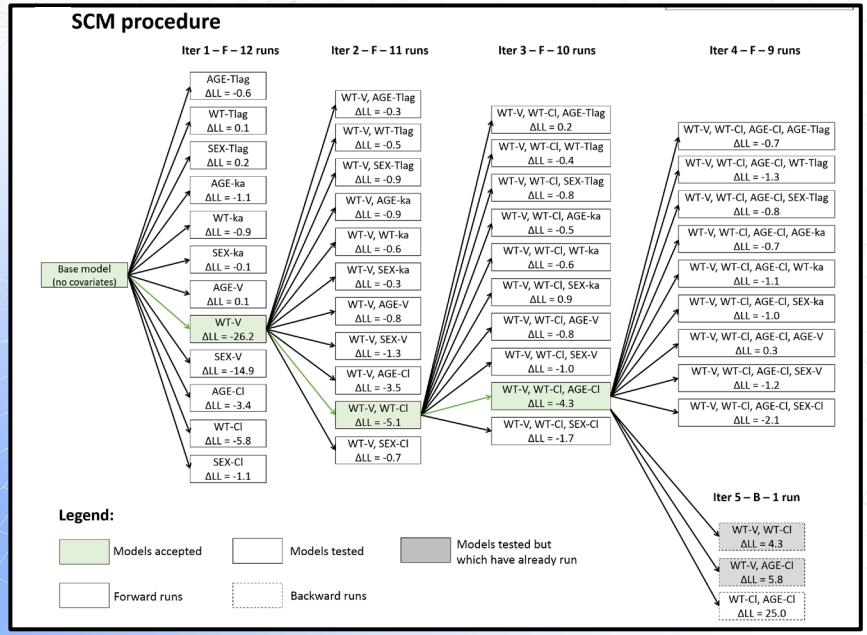


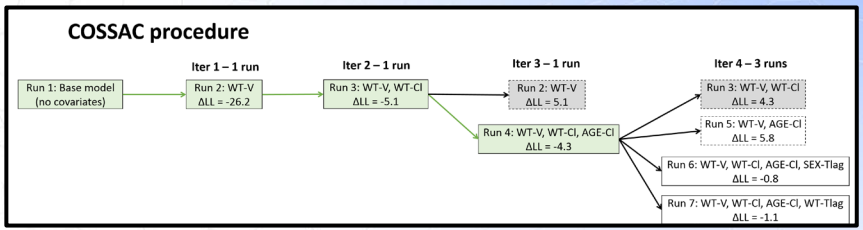
Ayral (Cellière), G. et al, *CPT Pharmacometrics Syst. Pharmacol.* (2021)

Number of runs

SCM: 43 runs
runs ≈ #par x #cov x #relations

COSSAC: 7 runs
runs ≈ #par x #cov





Ayral (Cellière), G. et al, *CPT Pharmacometrics Syst. Pharmacol.* (2021)

COSSAC: comparison to SCM on real datasets

Data set	Characteristics	Parameters (*italic: no variability*)	Covariates	COSSAC No. runs	COSSAC Final model	SCM No. runs	SCM Final model	ΔLL	ΔBICc	Ratio # runs
Remifentanil PK	Linear PK - SD - dense 65 indiv - 1992 obs	6 - Cl, V1, Q2, V2, Q3, V3	6 - SEX, logAGE, logBSA, logHT, logLBM, logWT	13	SEX – **V3** logAGE - Cl, Q2, V2, V3 logBSA - Cl logLBM – V1, **V2**	295	logAGE – Cl, Q2, **Q3**, V2, V3 logBSA - Cl logHT – **V2** logLBM – V1	−3.8	0.4	22.7
Theophylline PK	Linear PK - SD - dense 12 indiv - 120 obs	3 - ka, V, Cl	2 - SEX, logWT	6	None	7	None	Identical		1.2
Verapamil PK	Linear PK - SD - dense 22 indiv - 330 obs	6 - Tlag, ka, Cl, V1, Q, V2	7 - SEX, RACE, logAGE, logHT, logWT, logDIABP, logSYSBP	34	SEX - Cl, **V1**, *ka* logAGE - *ka* logWT - Q, V2	241	SEX - Cl, *ka* logAGE - *ka* logWT - Q, V2	−2.6	0.5	7.1
GBR12909 PK	Linear PK - MD - dense 12 indiv - 232 obs	5 - ka, V, k, k12, k21	2 - SEX, logWT	5	logWT - *ka*	20	logWT - *ka*	Identical		4
Quinidine PK	Linear PK - SD - dense 21 indiv - 315 obs	6 - Tlag, ka, Cl, V1, Q, *V2*	7 - SEX, RACE, logAGE, logHT, logWT, logDIABP, logSYSBP	20	SEX - Cl, V1	124	SEX - Cl, V1	Identical		6.2
Quinidine sparse PK	Linear PK - MD - sparse 136 indiv - 361 obs	3 - ka, V, Cl	7 - RACE, HEART, ETHANOL, SMOKE, logAGE, logHT, logWT	11	None	22	None	Identical		2
Tobramycin sp. PK	Linear PK - MD - sparse 97 indiv - 322 obs	2 - V, Cl	4 - SEX, logAGE, logCLCR, logWT	7	logCLCR - Cl logWT - V	22	logCLCR - Cl logWT - V	Identical		3.1
Cisplatine PK	Linear PK - MD - dense 23 indiv - 524 obs	6 - Cl, V1, Q2, V2, *Q3*, *V3*	5 - SEX, logAGE, logBSA, logHT, logWT	16	logBSA - V1	60	logBSA - V1	Identical		3.8
Theophylline ER PK	Linear PK - SD - dense 18 indiv - 362 obs	7 - ka1, ka2, F1, Tlag1, diffTlag2, V, Cl	3 - logAGE, logHT, logWT	17	logWT - Tlag1, **V**	61	logAGE – **ka2** logWT – Tlag1	0.5	0.5	3.6
IgG1 mAb PK	TMDD PK - SD - dense 28 indiv - 263 obs	7 - V, *kint*, kon, R0, Cl, Q, V2	2 - RA, logWT	13	RA - Cl, V, kon logWT - V	63	RA - Cl, V, kon logWT - V	Identical		4.8
Remifentanil seqPD	PD - SD - dense 61 indiv - 3989 obs	5 - ke0, E0, I_{max}, IC_{50}, *gam* (indiv. PK param fixed)	6 - SEX, logAGE, logBSA, logHT, logLBM, logWT	29	SEX - gam logAGE - IC_{50}, gam, ke0 logHT - gam	194	SEX - gam logAGE - **E0**, IC_{50}, gam, ke0 logHT - gam	8.4	4.3	6.7

Ayral (Cellière), G. et al, *CPT Pharmacometrics Syst. Pharmacol.* (2021)

COSSAC: comparison to SCM on real datasets

Data set	Characteristics	Parameters (*italic: no variability*)	Covariates	COSSAC No. runs	COSSAC Final model	SCM No. runs	SCM Final model	ΔLL	ΔBICc	Ratio # runs
Dofetilide PK/PD	Joint PK/PD - SD - dense 22 indiv - 328x2 obs	8 - Tlag, ka, Cl, V1, *Q*, *V2*, *intercept*, slope	7 - SEX, RACE, logAGE, logHT, logWT, logDIABP, logSYSBP	60	RACE - *intercept*, **ka** logWT - V1	220	RACE - *intercept* logSYSBP - ***intercept*** logWT - V1	0.5	0.2	3.7
MIDD (ASCPT Gran Prix)	Joint model parent/ metab/urine/PD 176 indiv – 2664 + 2723+ 147 + 2600 obs	9 - ka, V, Cl, Clr, Clm, Vm, R0, kdeg, IC_{50}	8 - SEX, ESRD, logAGE, logHT, logWT, logALB, nDiseases, nDrugs	20	ESRD - **Cl** logAGE - Cl logALB - Cl, **Clr**, V, ka logWT - **Clr**	421	logAGE - Cl logALB - Cl, V, ka nDiseases - $\mathbf{k_{deg}}$	−40	−29	21.1
Warfarin PK/PD	Joint PK/PD - SD - dense 32 indiv - 247 + 232 obs	8 - Tlag, ka, V, Cl, R0, kout, I_{max}, IC_{50}	3 - SEX, logAGE, logWT	11	logWT - V	48	logWT - V	Identical		4.4
Cholesterol	Disease progression 200 indiv - 1044 obs	2 - Chol0, slope	2 - SEX, logAGE	5	logAGE - Chol0, slope SEX - slope	12	logAGE - Chol0, slope SEX - slope	Identical		2.4
Alzheimer	Disease - count data 896 indiv - 3707 obs	2 - p0, slope	7 - SEX, RACE, APOE, logAGE, logBMI, logHT, logWT	8	APOE - alpha, p0 logAGE - alpha, p0 logBMI - alpha logWT - p0	82	APOE - alpha, p0 logAGE - alpha, p0 logBMI - alpha logWT - p0	Identical		10.3
Lung cancer survival	Time-to-event 228 indiv - 165 events	2 - Te, *k*	5 - SEX, ecogPH, karnoPAT, karnoPH, age	13	AGE - k ecogPH - Te SEX - Te	36	AGE - k ecogPH - Te SEX - Te	identical		2.8

Ayral (Cellière), G. et al, *CPT Pharmacometrics Syst. Pharmacol.* (2021)

Performance of COSSAC

- For the large majority of cases, the final covariate model is identical (11 out of 17) or very similar (4 out of 17) with COSSAC and SCM

	Number of models	Percentage
Identical models	11	64%
COSSAC model slightly better ($\Delta LL < 3.84$)	2	12%
SCM model slightly better ($\Delta LL < 3.84$)	2	12%
COSSAC model significantly better	1 ($\Delta LL = 40$)	6%
SCM model significantly better	1 ($\Delta LL = 8.4$)	6%

- COSSAC requires 2x - 20x fewer runs
- Makes covariate search possible for big models that are intractable with SCM

SAMBA

Stochastic **A**pproximation for **M**odel **B**uilding **A**lgorithm

Prague, M. & Lavielle, M. SAMBA: A novel method for fast automatic model building in nonlinear mixed-effects models. *CPT Pharmacometrics Syst. Pharmacol.* **11**, 161–172 (2022).

SAMBA key idea

Add several covariates at once to be even faster.

SAMBA key idea

Add several covariates at once to be even faster

First idea:
add all covariates
having a low p-value

Pearson's correlation test and/or ANOVA

eta_Tlag

	COEFF	STATISTICS	P-VALUE
sex		1.14	2.95e-1
age	0.075	0.41	6.84e-1
wt	0.0013	0.0074	9.94e-1

eta_ka

	COEFF	STATISTICS	P-VALUE
sex		0.036	8.5e-1
age	0.0043	0.024	9.81e-1
wt	0.24	1.33	1.94e-1

eta_V

	COEFF	STATISTICS	P-VALUE
sex		18.59	1.61e-4
age	-0.039	-0.21	8.32e-1
wt	0.75	6.24	7.22e-7

eta_Cl

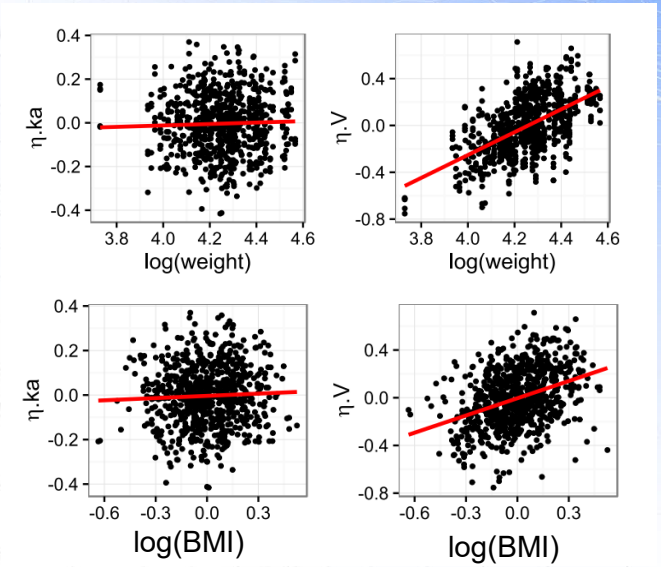
	COEFF	STATISTICS	P-VALUE
sex		0.38	5.43e-1
age	0.32	1.83	7.78e-2
wt	0.35	2.08	4.65e-2

SAMBA key idea

Add several covariates at once to be even faster

First idea:
add all covariates
having a low p-value

Problem:
some covariates might be
correlated with each other and
carry redundant information
e.g weight and BMI



SAMBA key idea

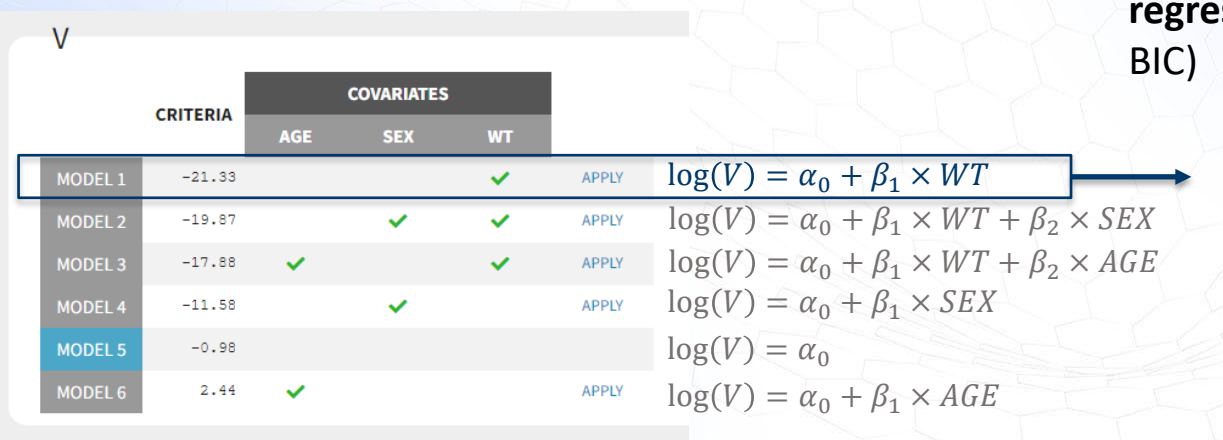
Add several covariates at once to be even faster

First idea:
add all covariates having a low p-value

Problem:
some covariates might be correlated with each other and carry redundant information
e.g weight and BMI

Solution:
For each parameter,

- **linear regression** between individual parameters and combinations of covariates
- Add covariates used in the **best regression model** (based on a BIC)



lowest BIC among regression models for V

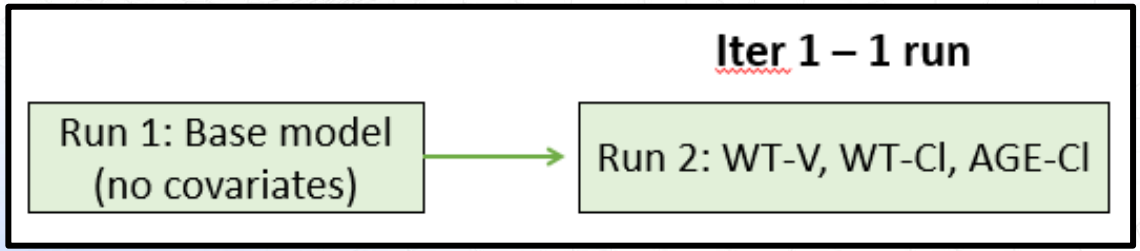
SAMBA procedure: overview

- For each parameter, try all possible linear regression models
- Add the covariates corresponding to the best (BIC) linear model and run
- Repeat if best linear model has changed, otherwise stop

Prague, M. & Lavielle, M. SAMBA: A novel method for fast automatic model building in nonlinear mixed-effects models. *CPT Pharmacometrics Syst. Pharmacol.* **11**, 161–172 (2022).

SAMBA procedure example

4 parameters: Tlag, ka, V, Cl and 3 covariates: Age, Weight, Sex

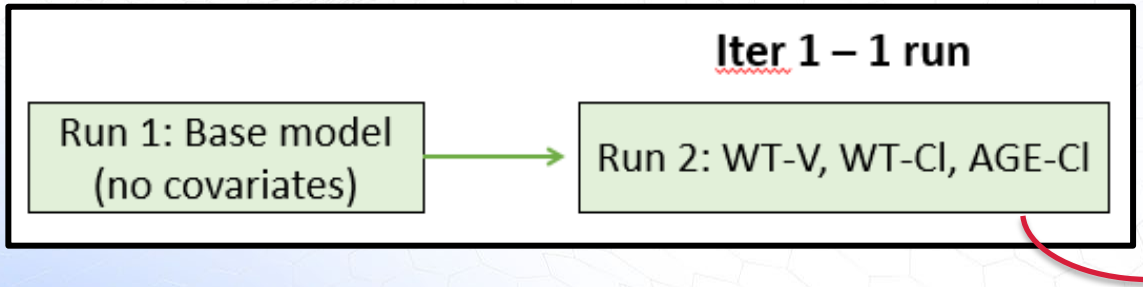


	COVARIATES		
	AGE	SEX	WT
Tlag			
ka			
V			✓
Cl	✓		✓

For each parameter, best linear regression between samples from conditional distribution and covariates

SAMBA procedure example

4 parameters: Tlag, ka, V, Cl and 3 covariates: Age, Weight, Sex



Best linear regression between samples from conditional distribution and covariates is the same => STOP

	COVARIATES		
	AGE	SEX	WT
Tlag			
ka			
V			✓
Cl	✓		✓

SAMBA: comparison on real datasets

		SCM		COSSAC		SAMBA		ΔBICc	
Dataset	**Characteristics**	**#Runs[b]**	**Final Model[a]**	**#Runs[b]**	**Final Model 1**	**#Runs[b]**	**Final Model[a]**	**SAMBA-SCM**	**SAMBA-COSSAC**
Warfarin	32 ind. - 247 obs.	44	logtWt - *V*, *Cl*	4	Identical	2	Identical	0	0
Linear PK	4 param. - 3 cov.		logtAge - *C*						
	4 re - 1 outcome								
Remifentanil	65 ind. - 1992 obs.	295	logLBM – *V*1	13	logLBM - *V*1, *V*2	4	logLBM - *V*1	0.8	0.5
Linear PK	6 param. - 6 cov.		logAGE - *Cl*, Q2, **Q3**, *V*2, *V*3		logAGE - *Cl*, *Q*2, *V*2, *V*3		logAGE - *Cl*, *Q*2, **Q3**, *V*2, *V*3		
	4 re - 1 outcome		logBSA - *Cl*		logBSA - *Cl*		logBSA - *Cl*		
			logHT - **V2**						
					SEX - **V3**		SEX - **V2**		
Theophylline	12 ind. - 20 obs.	12	logtWEIGHT - **ka**	4	Identical	2	Identical	0	0
Linear PK	3 param. - 2 cov.								
	4 re - 1 outcome								
Quinidine	136 ind. - 361 obs.	22	none	11	Identical	1	Identical	0	0
Sparse PK	3 param. - 2 cov.								
	3 re - 1 outcome								
Tobramycin	97 ind. - 322 obs.	22	logCLCR - *Cl*	6	logCLCR - *Cl*	2	logCLCR - *Cl*	4.2	4.2
Sparse PK	3 param. - 2 cov.		logWT - **V**		logWT - **V**		logWT - **Cl**		
	2 re - 1 outcome								
Theophylline	18 ind. - 362 obs.	98	logWT - **Tlag1**, **V**	8	logWT - **Tlag1**	6	logWT - **F**, **V**	−11.7	−27
Ext. Rel.	7 param. - 3 cov.				logAGE -**ka2**		logAGE -**F**		
Linear PK	7 re - 1 outcome						logHT **ka1**, **ka2**, **Tlag1**, **diffTlag2**		

Prague, M. & Lavielle, *CPT Pharmacometrics Syst. Pharmacol* (2022).

SAMBA: comparison on real datasets

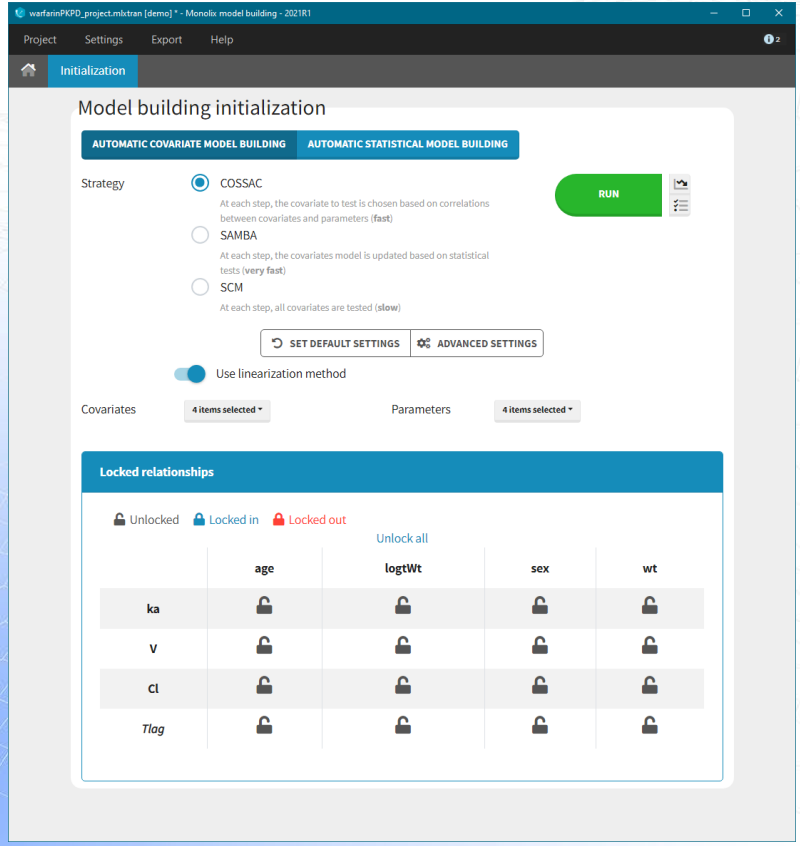
		SCM		COSSAC		SAMBA		ΔBICc	
Dataset	**Characteristics**	**#Runs[b]**	**Final Model[a]**	**#Runs[b]**	**Final Model 1**	**#Runs[b]**	**Final Model[a]**	**SAMBA-SCM**	**SAMBA-COSSAC**
Warfarin	32 ind. - 247+232 obs.	92	logWT - *Cl*	10	logWT - *Cl*	2	logWT - *Cl*, **V**	−1.4	−1.4
PK/PD	8 param. - 3 cov.						logAGE -**Cl**, **R0**		
Joint	8 re - 2 outcomes								
Cholesterol	200 ind. - 1044 obs.	12	logAGE - *Chol0*, **slope**	5	logAGE - *Chol0*, **slope**	2	logAGE - *Chol0*	13.5	13.5
Disease	2 param. - 2 cov.		SEX - **slope**		SEX - **slope**				
Progression	2 re - 1 outcome								
Alzheimer	896 ind. - 3707 obs.	73	APOE - *alpha*, *p0*	8	APOE - *alpha*, *p0*	2	APOE - *alpha*, *p0*	6	1.5
Sparse PK	2 param. - 7 cov.		logAGE - *p0*, **alpha**		logAGE - *p0*, **alpha**		logAGE - *p0*		
	2 re - 1 outcome		logBMI - **alpha**		logBMI -**alpha**				
			logWT - *p0*		logWT - *p0*		logWT - *p0*		
Tranexamic	166 ind. - 817 obs.	298	GROUP - *Cl*, *V2*	12	Identical	2	Identical	0	0
PK	4 param. - 10 cov.		logBMI - Cl						
	4 re - 1 outcome		logCOCK - Cl						
			logLBW - Q						
			logWeight - V2						

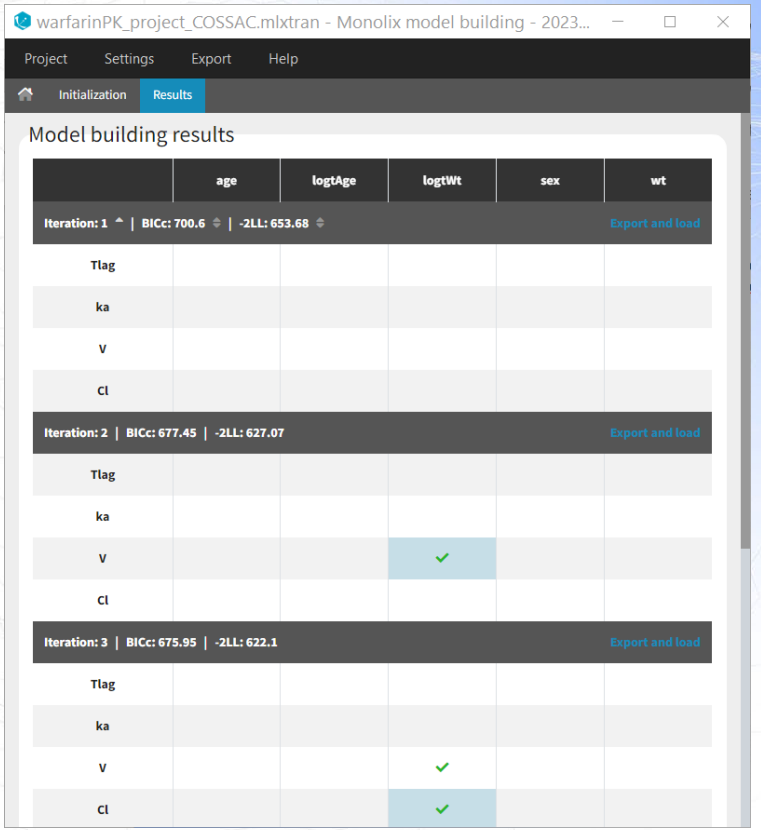
Prague, M. & Lavielle, *CPT Pharmacometrics Syst. Pharmacol* (2022).

Performance of SAMBA

- For the majority of cases (7 out of 10), the final covariate model is identical (4 cases) or very similar (3 cases) with SAMBA and SCM
- SAMBA requires only very few runs (usually 2-3, max 6)
- Makes covariate search possible for big models that are intractable with SCM or COSSAC

SAMBA and COSSAC are available in Monolix





For versions >= 2019R1

Publications using COSSAC

- COSSAC is available in Monolix since 2019, and article was published in 2021
- Method was used and cited in 22 scientific publications since then
- A few examples:
  - **Idorsia**: Krause, A., Lott, D., Brussee, J. M., Muehlan, C., & Dingemanse, J. (2023). Population pharmacokinetic modeling of **daridorexant**, a novel dual orexin receptor antagonist. CPT: Pharmacometrics & Systems Pharmacology, 12(1), 74–86.
  - **Sanofi**: Thai, H.-T., Gaudel, N., Cerou, M., Ayral, G., Fau, J.-B., Sebastien, B., van de Velde, H., Semiond, D., & Veyrat-Follet, C. (2022). Joint modelling and simulation of M-protein dynamics and progression-free survival for alternative **isatuximab** dosing with pomalidomide/dexamethasone. British Journal of Clinical Pharmacology, 88(5), 2052–2064.
  - **BMS**: Cheng, Y. et al. Model-based analysis for the population pharmacokinetics of **iberdomide** and its major active metabolite in healthy subjects and patients with relapsed and refractory multiple myeloma. Br. J. Clin. Pharmacol. 89, 316–329 (2023).
  - **GSK:** Yang, S., Simeoni, M. & Beerahee, M. Longitudinal Model-Based **Meta-Analysis** of Lung Function Response to Support Phase III Study Design in Chinese Patients With **Asthma**. Clin. Pharmacol. Ther. 111, 1286–1295 (2022).
  - **Academics:** Suñer, C. et al. **Viral dynamics** in patients with monkeypox infection: a prospective cohort study in Spain. Lancet Infect. Dis. 23, 445–453 (2023).

Acknowledgements

Marc Lavielle and **Mélanie Prague** (Inria)

Key ideas of COSSAC and SAMBA
Implementation of SAMBA in Rsmlx

Géraldine Cellière, Jonathan Chauvin, Jean-François Si Abdallah (Simulations Plus)

Key ideas of COSSAC and SAMBA
Implementation of COSSAC in Rsmlx
Optimization of COSSAC and SAMBA in Monolix

References

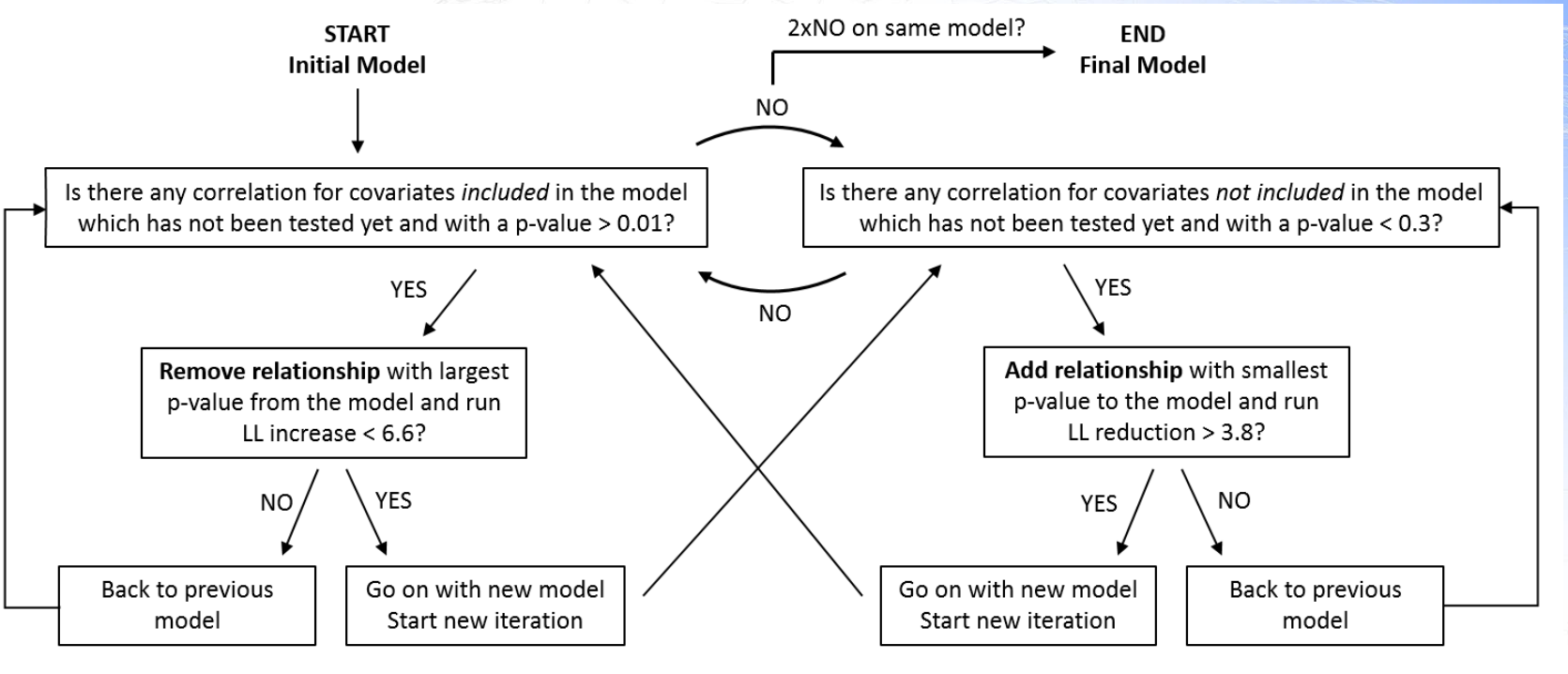
- Sampling from the conditional distribution
  Lavielle, M. & Ribba, B. Enhanced Method for Diagnosing Pharmacometric Models: Random Sampling from Conditional Distributions. *Pharm. Res.* (2016)

- COSSAC:
  Ayral (Cellière), G., Si Abdallah, J. F., Magnard, C. & Chauvin, J. A novel method based on unbiased correlations tests for covariate selection in nonlinear mixed effects models: The COSSAC approach. *CPT Pharmacometrics Syst. Pharmacol.* **10**, 318–329 (2021).

- SAMBA:
  Prague, M. & Lavielle, M. SAMBA: A novel method for fast automatic model building in nonlinear mixed-effects models. *CPT Pharmacometrics Syst. Pharmacol.* **11**, 161–172 (2022).

Additional slides in case of questions

COSSAC detailed procedure



Ayral (Cellière), G. et al, *CPT Pharmacometrics Syst. Pharmacol.* (2021)

Availability in Monolix: scripts/batch

- In R via lixoftConnectors package
  - calls the same C++ code than the GUI
  - `runModelBuilding(covariates= …, parameters=…, strategy=…, criterion=…, relationships=…, threshold=…)`

- In R via the Rsmlx package
  - beta implementation, slightly different from the final one in Monolix
  - source code is open
  - `covariateSearch(project=…, method=…, covToTest=…, paramToUse=…, testRelations=…, settings=…)`

- In command line
  - `monolix.bat --no-gui -t modelBuilding -s cossac -a 0.06 -r 0.001 –p "C:\Users\celliere\lixoft\monolix\demos\example.mlxtran"`

Performance of SAMBA in a simulation study

TABLE 2 Performance of the SAMBA algorithm for the selection of the covariate model in a simulation study using a one-compartment PK model

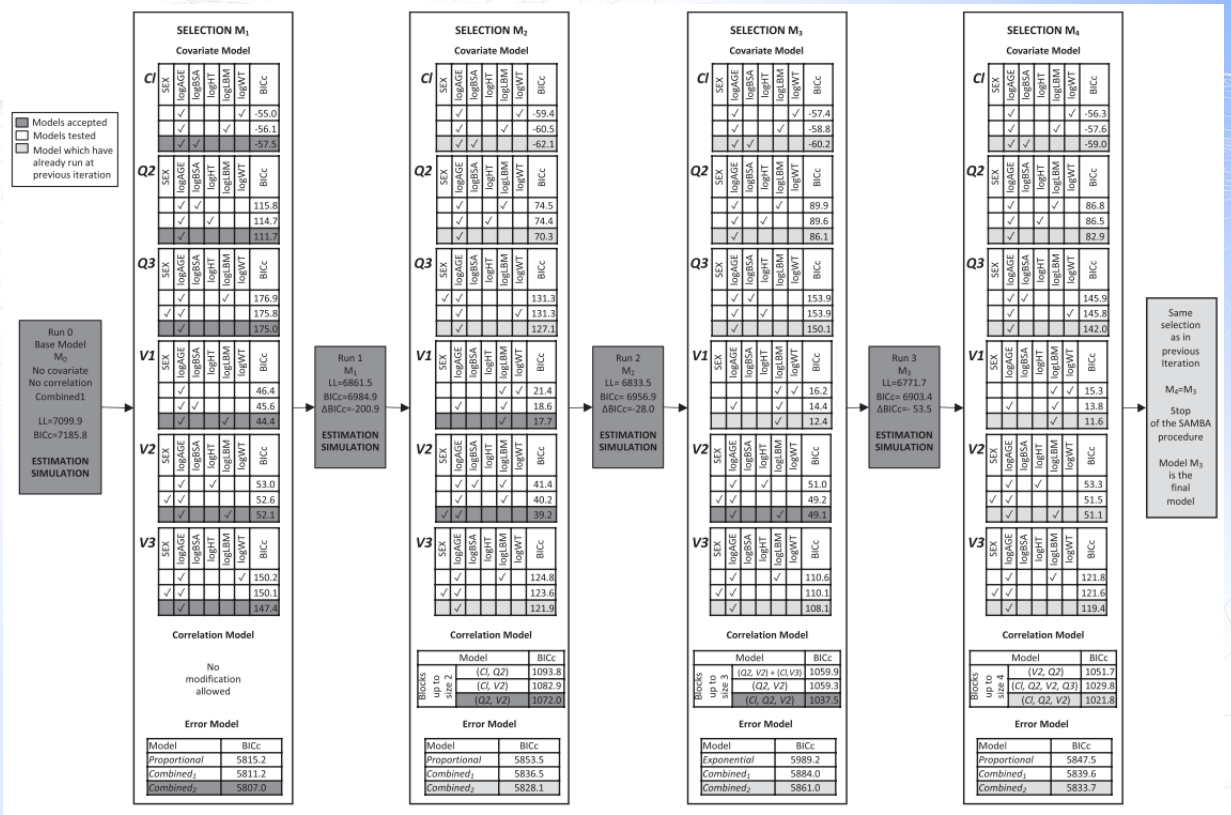
	Rsmlx			Monolix		
Covariates	***ka***	***V***	***Cl***	***ka***	***V***	***Cl***
C_1	2	100	100	2	100	100
C_2	0	1	100	0	1	100
C_3	1	2	1	2	2	1
C_4	0	3	4	0	3	4
C_5	0	1	1	1	2	1

One hundred datasets of 100 individuals with 11 observations each have been generated. True model $\mathscr{M}^*$ includes an effect of C_1 on V and Cl and an effect of C_2 on Cl. The percentages of times (over 100 replicates) each covariate-parameter relationship is selected in the final model are displayed. Implementation of SAMBA in *Rsmlx* and Monolix are compared.

Automatic model building: SAMBA

SAMBA can work on

- Covariates
- Correlations between random effects
- Error models



Calculating number of runs

The reduction in the number of runs between SCM and COSSAC will depend on the number of parameters p, the number of covariates c and the number of parameter-covariate relationships r in the final model.

In a SCM procedure, the first forward step requires $p \times c$ runs, the second $(p \times c) - 1$ runs, etc, until r relationships have been added. In the final forward step, $(p \times c) - r$ runs are done and none is accepted. In the backward step(s), the r relationships can be tested for removal, corresponding to r runs. There is in general only one backward step. We thus obtain the following approximate formula:

$$\#\text{runs}_{\text{SCM}} \approx \left(\sum_{i=0}^{r} p \times c - i\right) + r$$

$$\#\text{runs}_{\text{SCM}} \approx p \times c \times (r+1) - r \times (r+1)/2 \; + r$$

With a COSSAC procedure, most often the first forward steps lead to an accepted model until the final model is reached. This corresponds to r forward runs approximately. They are then followed by forward not accepted models. Once models tend not be accepted anymore, there remain $(p \times c) - r$ models which can be tested for additions, corresponding to at maximum $(p \times c) - r$ runs. In practice only a fraction has a sufficiently low p-values to be run and evaluated. The backward steps are mostly models which have already run and do not require a new run, but we will consider that r relationships can be tested for removal. This leads to the approximate formula:

$$\#\text{runs}_{\text{COSSAC}} \approx r + (p \times c) - r + r = r + (p \times c)$$

COSSAC vs SCM: different final models

For one run (remifentanil seqPD), the SCM method finds a model with one additional relationship (logAGE on E0) compared to COSSAC, which leads to an 8.4 points better LL. This better model is not tested by COSSAC because logAGE on E0 improves the LL only once covariates have been added on the gamma parameter, which has no variability and is tested after all others only. Running COSSAC again at the end would resolve the discrepancy but comes at a substantial cost in terms of runs. A similar situation happens for the Verapamil PK example.

On the opposite, for the model-informed drug development (MIDD) dataset, COSSAC finds a model that is 40 points better than SCM. The path of accepted runs taken by both methods is the same for the four first covariate additions. For the fifth, SCM adds nDiseases on the first order degradation rate (k_{deg}) (largest LL decrease) and no further addition leads to a sufficient LL improvement. COSSAC adds logWT on Clr (smallest correlation p value). The LL improvement of this addition is smaller than that of nDiseases on k_{deg}, but this turns to be an advantage afterward, as the end-stage renal disease (ESRD) on Cl and logALB on renal clearance (Clr) can be added as additional significant covariates.